

code009

July 2, 2024

1 Medidas de dispersión

Son el complemento perfecto para las medidas de tendencia central, pongamos de contexto un salón de clases en donde los estudiantes tienen diferentes edades, decimos **el promedio de edad es 7 años**, ese número me servía para decir, ya tengo idea que en ese salón los estudiantes son jóvenes, pero ¿qué tanto difieren esas edades respecto del valor central? Puede que haya estudiantes mas grandes o muy jóvenes 6 años, entonces necesito tener una noción matemática de como medir esa dispersion de los datos, eso lo podemos hacer con lo que estamos definiendo de la clase .

1.1 Dispersión en una distribución

- Rango
- Rango intercuartil
- Desviación estándar.

Tenemos un histograma y podemos ver que tiene una forma de curva de campana y esa forma es porque los datos tienden a seguir distribuciones de ese estilo, a eso se le asocia una distribución muy famosa que es **Distribución Gaussiana o Distribución normal o Distribución Estándar**.

Rango

El rango se mide restando el valor máximo con el valor mínimo, dando como resultado el rango de dispersión de la medida central.

$$R(x) = \max(x) - \min(x)$$

Rango intercuartil

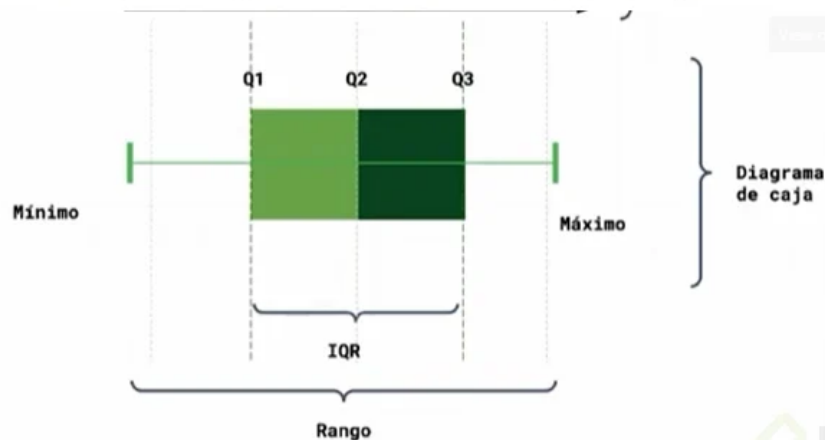
Cuando trabajamos con cuartiles, trabajamos con las cuartas partes de un rango de valores, donde tenemos Q_1, Q_2, Q_3

Entonces tenemos que:

- $Q_2 = \text{median}(x)$ esto quiere decir que siempre será la mitad (50%)
- $Q_1 = \frac{\text{median}(x) + \min(x)}{2}$ esto quiere decir que, será el valor que se encuentra entre el mínimo y la mitad (25%)
- $Q_3 = \frac{\text{median}(x) + \max(x)}{2}$ esto quiere decir que, será el valor que se encuentra el máximo y la mitad (75%)

A estos son los que denominaremos, rangos intercuartiles.

La mejor forma de graficar rangos intercuartiles, es usar un diagrama de caja



1.1.1 Rango

La manera mas rápida de medir la dispersion de datos es; **tomar el valor máximo y el valor mínimo** y hacer una resta de ellos, al hacerlo tengo la maxima medida de dispersión de ese conjunto. En los ejemplos anteriores es ese corchete que abarca la distancia del valor mínimo y el valor máximo entre los datos.

1.1.2 Rango intercuartil

Tiene que medir con relación a algo que se llama **cuartiles** y tienen como propósito dividir a mi rango de datos en 4 partes homogéneas (quiere decir que tengo la misma cantidad de datos en cada subdivision) 4 subdivisiones. La mediana Q_2 parte a la mitad un conjunto de datos. - 1er cuartil es el que esta a la mitad entre mediana y el inicio de los datos. - 2do cuartil es la mediana. - 3er

cuartil es la mitad entre la mediana y el valor máximo

La distancia entre el Q1 y el Q3 es lo que llamaremos el **Rango Intercuartil (IQR)**. IQR es el nombre en Ingles de **Rango intercuartil**.

Rango

El rango se mide restando el valor máximo con el valor mínimo, dando como resultado el rango de dispersión de la medida central.

$$R(x) = \max(x) - \min(x)$$

Rango intercuartil

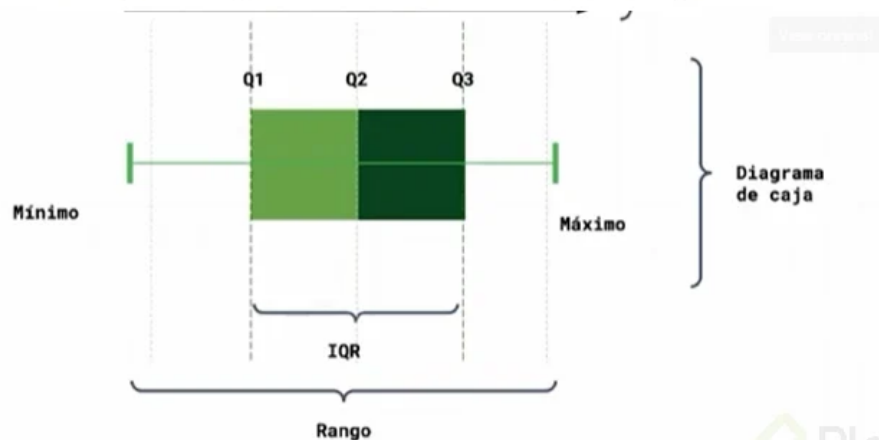
Cuando trabajamos con cuartiles, trabajamos con las cuartas partes de un rango de valores, donde tenemos Q_1, Q_2, Q_3

Entonces tenemos que:

- $Q_2 = \text{median}(x)$ esto quiere decir que siempre será la mitad (50%)
- $Q_1 = \frac{\text{median}(x) + \min(x)}{2}$ esto quiere decir que, será el valor que se encuentra entre el mínimo y la mitad (25%)
- $Q_3 = \frac{\text{median}(x) + \max(x)}{2}$ esto quiere decir que, será el valor que se encuentra el máximo y la mitad (75%)

A estos son los que denominaremos, rangos intercuartiles.

La mejor forma de graficar rangos intercuartiles, es usar un diagrama de caja



El anterior Diagrama se llama **Diagrama de caja o Box plot**, es la visualización por excelencia de la forma en que está ubicada la dispersion de los datos, en referencia a la mediana en particular.

1.2 Extras

Rango

El rango se mide restando el valor máximo con el valor mínimo, dando como resultado el rango de dispersión de la medida central.

$$R(x) = \max(x) - \min(x)$$

Rango intercuartil

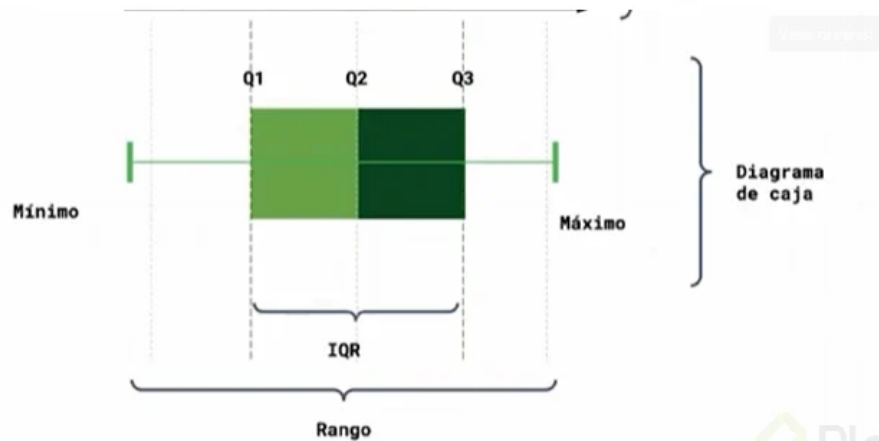
Cuando trabajamos con cuartiles, trabajamos con las cuartas partes de un rango de valores, donde tenemos Q_1, Q_2, Q_3

Entonces tenemos que:

- $Q_2 = \text{median}(x)$ esto quiere decir que siempre será la mitad (50%)
- $Q_1 = \frac{\text{median}(x) + \min(x)}{2}$ esto quiere decir que, será el valor que se encuentra entre el mínimo y la mitad (25%)
- $Q_3 = \frac{\text{median}(x) + \max(x)}{2}$ esto quiere decir que, será el valor que se encuentra el máximo y la mitad (75%)

A estos son los que denominaremos, rangos intercuartiles.

La mejor forma de graficar rangos intercuartiles, es usar un diagrama de caja



- ¿Cómo calcular el IQR